



TITLE:

分子進化の中立説と集団遺伝学(数学者のための分子生物学入門,研究会報告)

AUTHOR(S):

高畑, 尚之

CITATION:

高畑, 尚之. 分子進化の中立説と集団遺伝学(数学者のための分子生物学入門,研究会報告). 物性研究 2003, 81(1): 60-68

ISSUE DATE:

2003-10-20

URL:

<http://hdl.handle.net/2433/97614>

RIGHT:

分子進化の中立説と集団遺伝学 総合研究大学院先導科学研究科・高畑 尚之

§ 歴史

1859年、イギリスの生物学者、ダーウィンが出版した「種の起原」により進化論は世界中に広まった。しかし、進化の事実を主張したのはダーウィンが最初ではない。ラマルクは19世紀初頭に生物が下等から高等へ進化してくことを説いた。しかし、彼の説は獲得形質の遺伝に基礎をおくものだったので、遺伝学の発達により否定された。

19世紀半ばになるとダーウィンやウォレスにより進化が定式化された。ダーウィンは進化を変化の伴う由来であると定義し、種の起原をまとめた。ダーウィンが種の起原をまとめるにあたって、彼を悩ました最大の難点は遺伝の仕組みが分からないことだった。同時期オーストリアの修道院ではメンデルがエンドウを用いて交配実験を行っていた。メンデルは1865年にその結果を発表したが、長い間世に認められず、やっと1900年になって再発見され注目を浴びるようになった。つまり、ダーウィンはメンデルの遺伝の法則を知らなかったのである。ダーウィンは遺伝の本体をあいまいのまま、進化とは親、子、孫と代々伝達するもので、それを伝える物質は細胞の中にあることを見抜いた。ダーウィンが主張した最も重要なことは、個体が集団をなし、この集団に変異が起きて、代々伝達するということである。伝達する過程で変異が起きた個体が増えるか、減るかは分からないが、ここに発展的とか向上的という価値観は存在しない。伝達する過程を決めているメカニズムは集団レベルで存在し、そのうちの1つが確率過程である。この時代の生物学は分野的な分類はなく、非常に総合的なものだった。

その中からモルガン達一部が発生機構論を樹立した。その後、イギリスのウェルドンとピアソンは遺伝現象の統計学的分析に関心を持った。1920年代になると、ダーウィンの進化理論をメンデルの遺伝学に基づき生物統計学的に基礎づけようとする努力が実り、集団遺伝学が誕生した。そして、数学的議論とデータを結びつける分子進化学が発達した。この分野には中立説を唱えた木村資生氏がいる。

ダーウィンの進化論の弱点の1つは遺伝の仕組みが不明なことであった。メンデルの遺伝学はこの弱点を補うはずであったが、1900年、メンデルの法則が再発見された当初は対立するものだった。1930年以降、進化のメカニズムとしての突然変異や組換えなど生物学の幅広い領域の成果を取り入れ、遺伝学と自然淘汰説によって進化を理解すべきだという考えが生まれた。この研究は進化の総合説と呼ばれる。現在では、遺伝子レベルの解析が

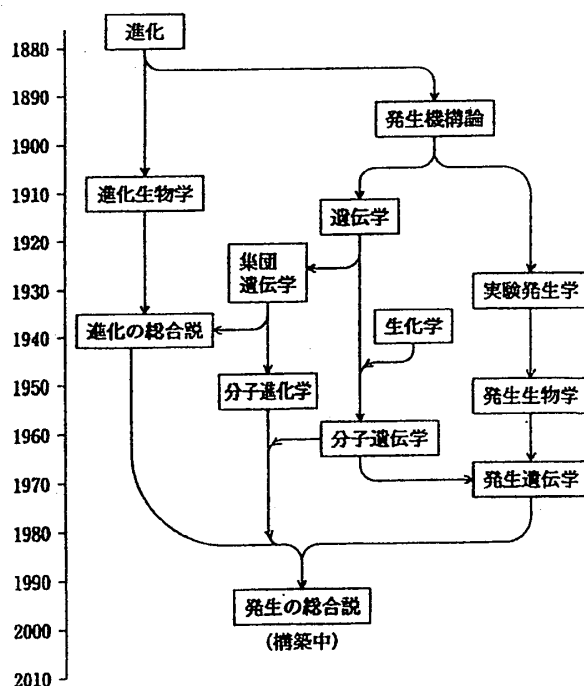


図1 進化遺伝学分野の分裂を統合

可能となり、ゲノムレベルでの情報を取り入れることによって発生学との総合説を構築中である。

§ 分子進化の機構

進化とは集団全体の変化であり、個々の DNA ではない。集団に変異が生じ、少なくとも 2 つ以上のタイプが共存し、この状態から変異がだんだん集団に広まるか、消滅するかどちらかの状態になる。ほとんどの変異は消滅してしまうが、上手くいけば変異が集団全体に広がる。集団全体に変異が広まった状態のことを固定という。進化にとって重要なことは DNA を一斉にある方向へ変えるメカニズムは存在しないことである。

進化のメカニズムは 2 つに分けることが出来る。1 つは DNA の変異を生成するメカニズムで個体レベルの話である。もう 1 つは束となって存在する DNA 内にどのように変異が広まるかに関するメカニズムであり、こちらは集団レベルの話である。どちらの現象も確率過程である。

染色した核を光学顕微鏡で観察すると染色体が見える。光学顕微鏡は $0.2 \mu\text{m}$ までしか観察できないため、これ以上詳細に観察することはできないが、電子顕微鏡が開発されると 0.2nm の世界まで観察できるようになった。1953 年ワトソンとクリックが X 線による解析で DNA の二重らせん構造を解明した。DNA は塩基性なので酸性のタンパク質が結合してヌクレオソームという構造を取っている。さらにこれらはねじれて高次の DNA 構造（スーパーコイル）をとっている。ヒトはこれが 23 対あり、 $5 \mu\text{m}$ の細胞核の中に入っている。DNA は次世代に伝えられる時、スーパーコイルを解きほぐし、自分と同じコピーを作らなければならない。

遺伝子は DNA の至るところにあるのではなく、遺伝子として働く部分は DNA のうちほんの数%にすぎない。残りの大部分ははっきりした機能を持ち合わせてない。ヒトの場合は 36 億の塩基対に対して全遺伝子領域は 1.5 % である。

種	C value(Mb)	種	C value(Mb)
パン酵母	12	ヒト	3600
キノコ	17	タバコ	3800
海綿	54	イナゴ、バッタ	6600
ホヤ	160	ゾウリムシ	8600
イネ	590	タマネギ	15000
ウニ	870	ヒマラヤ杉	19000
ニワトリ	1200	イボイモリ	19000
コイ	1700	ユリ	36000
ヘビ	2100	赤松	68000
サメ	2700	シダ	160000
イヌ	2900	アメーバ	690000

図2 真核生物のゲノムサイズ

最近ではいろいろな生物のゲノム DNA 配列を決める計画が進んでいる。C-value とはゲノムサイズを表わしているが、C-value は種によって一定である (図 2)。ヒトはゲノムサイズが大きいからヒトになったわけでもなく、ゲノムサイズが小さいからヒトになったわけでもない。つまり、進化とゲノムサイズはほとんど無関係である。シダやアメーバーは非常に大きいゲノムサイズを持っている。ゲノムは普通、複製により 2 倍になるが、細胞分裂により半分になる。しかし、アメーバーなどは複製により 2 倍にはなるが、細胞分裂しないため 1 つの細胞当たりのゲノムサイズが大きくなるのである。

§DNA レベルの機構

DNA レベルの変異生成の機構は塩基の置換、組換え、重複、欠失である。塩基の置換は進化のマーカーであり、分子時計と密接に関係している。変化には 2 種類あり、1 つはピリミジン (T,C) とピリミジンとの間での置換、又はプリン (A,G) とプリンとの間での置換で transition と呼ばれている。もう 1 つはピリミジンとプリンとの置換で transversion と呼ばれる。transversion は transition に比べると起きにくい。また、タンパク質をコードしている所の変化にも 2 種類ある。1 つは同義置換といい AAT(Asn) が AAC(Asn) になるように塩基が変わってもアミノ酸は変わらない変化。もう 1 つは非同義置換といい、TTG(Leu) が ATG(Met) になるように塩基が変わることにより、アミノ酸まで変わってしまう変化である。これはコドンとアミノ酸が一对一の対応ではなく、一つのアミノ酸に対して二つ以上のコドンが対応しているためである。同義置換は非同義置換に比べると起きやすい。

組換えは非常に重要なメカニズムであるが、高等生物ではそのメカニズムはほとんどかかっていない。しかし、組換えの結果についてはすでに明らかになっている。もともとの DNA の束の中に変異がなければ組換えは同じ塩基配列間で行われるので組換えは意味がないが、変異がある状態でさらに組換えが起こると高い多様性が生じ、非常に大切なメカニズムとなる。gene conversion は組換えと似たメカニズムと考えることが出来る。現象的には組換えは 2 本の塩基配列に平等に交叉が起こってつなぎ合わさるが、gene conversion は一方が他方の配列に置き換わり、元の方は残ったままである。

ヒトとチンパンジーは塩基が 1~2% 違うだけである。わずかな塩基置換が起こっただけで形態はこんなに違うのは非常に不思議である。ただし、ゲノム全体を見ると DNA が重複していたり、欠損している所があり実際にはかなり違う。例えば、第 6 染色体を観察すると、ヒトにはある部分でチンパンジーにはない部分が 100kb ある。DNA の 98% 以上の部分は機能として働かないどうでも良い部分である。このような部分では塩基のコピーや削除が頻繁に起こる。このような事を起こす分子機構は unequal crossing over と言い、相同染色体が平等に並んで組換えが起こるのではなく、組換えが起こる前に不平等に並び、重複や欠失が起こる。

これは病気と深く関係していることがわかった。ハンチントン舞踏病は脳の神経細胞に変性が起こる病気であるが、この病気は CAG リピート異常によることが分かっている。グルタミンの合成を指示する塩基配列 CAG が正常の場合は 20 回くらい繰り返されているが、ハンチントン舞踏病の遺伝子では 100 回以上繰り返す。このように単純な塩基配列の重複や欠失が病気に関係している。

また、南極海の魚、ノトセニア（スズキ科）にもリピート配列が見られる。南極海に氷床が出来、寒くなり始めたのは今から2000万年前である。この時期までは生物の多様性が低かったが、そこに適応したのがノトセニアである。南極海は-1.9度なので、何もしなければ体液などは凍ってしまい、生存出来ない。そこで、この魚は不凍化タンパク質を獲得し、環境に適応していった。異なる遺伝子のつなぎ合わせを使って同じような配列を繰り返すことにより不凍化タンパク質を作り出している（図3）。このように特殊な環境では、場合によっては新しい遺伝子を生み出す。現在、南極海のバイオマスでは95%がこの魚であり、種の数でも50%を占めている。

抗凍結遺伝子

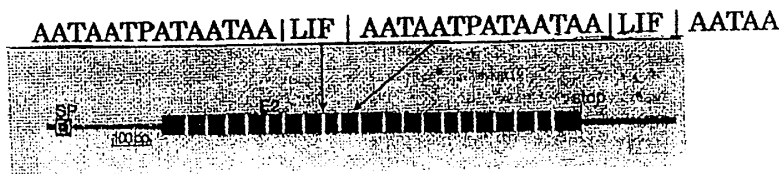


図3 南極海の魚と反復配列

§ 数学モデル

DNAの配列や、アミノ酸の配列を決めることは集団レベルでは容易ではなかったため、分子進化の研究はもともとタンパク質レベルで行われていた。タンパク質レベルの突然変異に関する数学モデルは以下のようなものがある。K個違った変異があると仮定する K-allele model。このモデルの1次元拡散方程式は $L = \frac{x(1-x)}{4N} \frac{d^2}{dx^2} + \{-\mu x + \mu'(1-x)\} \frac{d}{dx}$ である。ただし、 N は集団の個体数、 μ, μ' は突然変異率、 x は集団中に含まれる突然変異体の割合である。また、分子レベルでは新しい突然変異は新しいタイプを生成すると考えることにより無限にタイプがあると仮定する Infinitely many allele model。このモデルの拡散方程式は $L = \frac{x(1-x)}{4N} \frac{d^2}{dx^2} - \mu x \frac{d}{dx}$ である。当時はDNAのデータはなかったが、突然変異がDNAレベルで起こったとすると、それはいつも新しいDNA座に起こると仮定する Infinitely many site model が広く用いられている。そして、突然変異によって生じるアミノ酸の変化を電気泳動法により電荷の相違として観察し、突然変異が1つ起こるたびに状態が1つ動く Stepwise mutation model がある。しかし、このモデルは小さなりピート、特にDNAのマイクロサテライトにおいてはまだ確立されていない。タンパク質レベルの変異の研究は拡散方程式を利用し、集団中の変異がどの程度あるかということを調べるのが目的である。

DNAレベルでの数学モデルはA,T,C,Gの4つの塩基が自分以外のものにどのような割合で変化するかを考えるモデルで 4×4 行列の Markov Chain である。すべてが同等に変化すると仮定するのが Jukes-Cantor model であり、行列要素がすべて同じ値を持つ（図4）。Two parameter model は transition の割合と transversion の割合に変化をつけて考えるモデルである（図5）。12個のパラメーターの間にいろいろな条件をいれることにより、もっと一般的なモデルを考える事ができる。 4×4 行列では隣にどのような塩基が来るか

は無関係にならなければならないが、DNAは繋がっているので隣に影響されることもある。よって現在では16×16行列により隣の塩基の影響まで考慮している研究もある。

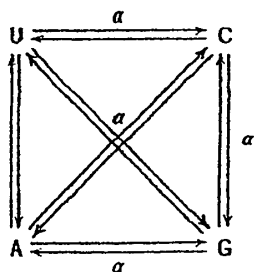


図4 Jukes-Cantor model

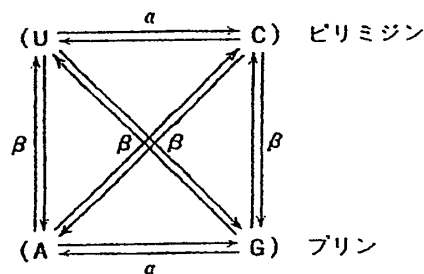


図5 Two parameter model

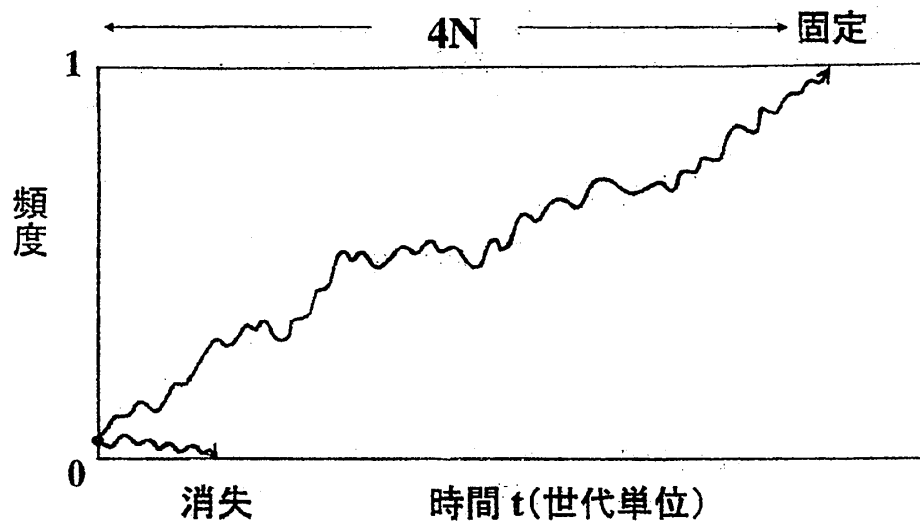
本質的に集団遺伝学は確率過程で考えなければならない。まず、ダーウィンの意味では自然淘汰に有利なものが出てきた時、どのように集団に広まるかを考える。集団内には異なった遺伝子型を持った個体が存在し、それらの間に生存力や適応度に差がある時、自然淘汰が働くという。適応度とは個体が次世代に子供をどれだけ残せるかを表わす量である。ここでは最も簡単な半数体モデルを考える。集団中に対立遺伝子 B と b があるとし、 B は野生型（正常）遺伝子とし、一方 b は突然変異によって生じた遺伝子とする。集団中の b の割合（即ち頻度）を p とすると B の頻度は $1-p$ である。適応度については B の淘汰値を1とした時、 b のそれは $1+s$ と表わされるとする。（この場合 s は b の有利さを表わし淘汰係数と呼ばれる。）次世代の b の頻度は $p' = \frac{(1+s)p}{1+sp}$ となるので、微分方程式で

は $\frac{dp}{dt} = sp(1-p)$ と近似される。ただし、この場合は半数体でありヒトのように父親と母親から遺伝子を1つずつ受け継ぎ2倍体になっている場合は適応度は両親からどのような遺伝子を受け継ぐかにより決まり、1つの遺伝子では決まらない。現実的にすればするほど差分方程式は複雑になるが、一般的に自然淘汰の数学モデルは非線形微分方程式で近似することができる。

進化が本質的に確率現象であるというのは、毎世代起こる遺伝子の無作為な抽出のためである。これを遺伝的浮動と呼ぶ。今、毎世代一定数 N の個体からなる2倍体の集団を考え、有性繁殖は次のような順序で行われるとする。まず、無限個の配偶子からなる遺伝子プールができ、このうちから N 個の雄性配偶子と N 個の雌性配偶子とが無作為に抽出され、その結合によって次代の N 個体が生成されるとする。この1世代当たりの変化量は2項分布 $Bin(2N, p)$ に従うことが分かる。

我々の興味あることは突然変異体がどのくらいの時間集団内に存続し、どのくらいの確率で消滅、又は固定するかということである。一般に集団内に出現した1個の突然変異遺伝子は大部分が集団から偶然に失われてしまう。運の良い場合は何世代も生き残り、ごく稀であるが集団全体に広がる。十分長い時間経った時、固定する確率は $\frac{1}{2N}$ である。また固定に要する時間は平均 $4N$ 世代であることが分かっている。（図6）

図6 遺伝的浮動による中立突然変異遺伝子の固定と消失の過程



進化を考える上で重要なことは”どのような交配をしているか?”ということである。最も簡単なモデルはでたために交配する任意交配集団である。これまで考えてきた N 個体の集団というのは、この任意交配集団を考えてきた。その他には Inbreeding と言い、近縁関係なほど交配しやすかったり、assortative mating といい、外見が似ているから交配しやすい、また disassortative mating といい自分と異なるタイプでなければ交配しないなど遺伝的な原因により任意交配からずれる交配構造や、今考えている集団の分集団がいくつもあり、地理的な隔離を伴う交配構造もある。地理的な隔離による交配構造には、移住個体がどこから来るかはまったくの偶然に依存し、遠くの集団も近くの集団も同じ割合で個体の交換が行われると考える Island model や、移住個体は隣接する集団から来ると考える stepping-stone model がある。Island model では2つの集団間の移住の効果は突然変異の効果と変わらない。

デモグラフィーに関する数学モデルには、新しい種が生まれる時、遺伝的浮動が強くなり、新しい遺伝的構成を持った集団が出来る Founder effect がある。また、Bottleneck effect は絶滅に瀕している種で見られるが、自分と良く似たタイプしか残っていないため非常に強い Inbreeding が起こっている。ヒトのように1万年位の間に1000万人から60億人に拡大したとき、どのような効果があるかを考える Population expansion がある。さらに、部分集団は常に安定して存在しているわけではなく、拡大したり、他のものに置き換わったりする。これに対応する考えは Extinction and re-colonization of local population である。

§ 分子進化の中立説

分子進化の2つの大きな特徴は年あたりの”速度の一定性”と、変化様式の”保守性”である。

一定性については、ヘモグロビンの α 鎖を例にとるのが良いだろう。脊椎動物に属する各種の生物のヘモグロビン α 鎖を比較し、種の分岐年代を考慮し、1年あたりアミノ酸座

位あたりの進化における置換率を推定してみると、ほぼ 10^{-9} という値になり、表現型レベルで急速に進化してきた生物群でも、何億年もほとんど形態的に変わっていない生物でも分子レベルでは進化の速度はほとんど同じという結果が得られる。(図7) 保守性の例で最も良く知られているのはプロインシュリンである。この分子は A,C,B から成るが、中央約 $\frac{1}{3}$ を閉める部分 C はインシュリンが出来る時に切り離され、A と B とが結合して活性のあるインシュリン分子となる。(図8) インシュリン A,B については、進化におけるアミノ酸置換率は1年あたり 0.4×10^{-9} と遅いが、切り離される C におけるアミノ酸置換率は 2.4×10^{-9} ときわめて速い。このことから遺伝子の重要な部分は変化しにくく、大事ではない部分の進化速度は速いこと、つまり変化が保守的であることが分かる。

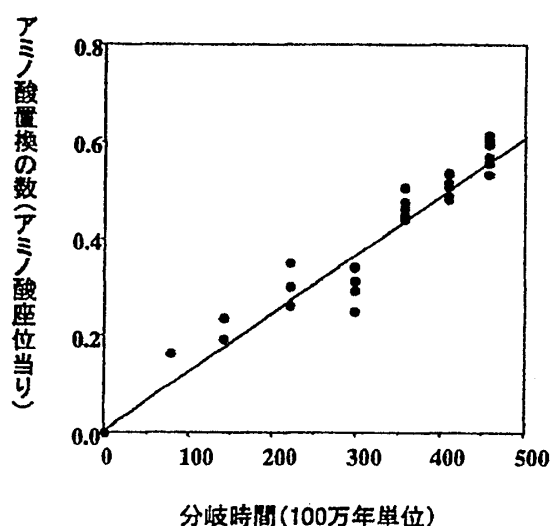


図7 ヘモグロビン α 鎖の分子進化時計

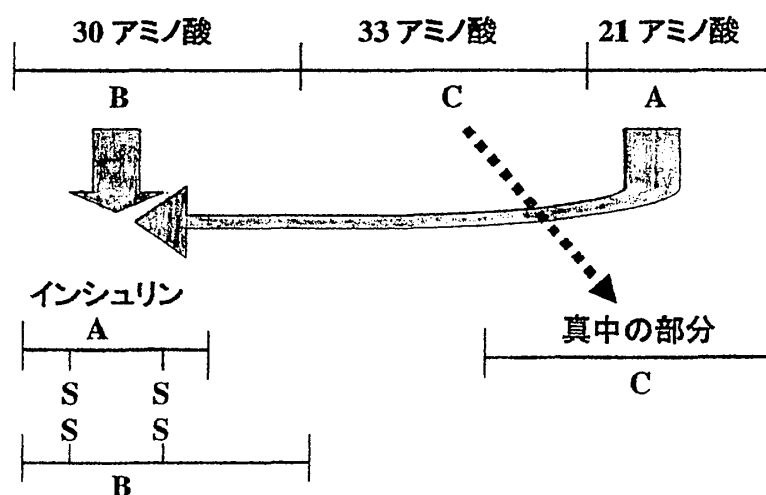


図8 プロインシュリンの進化速度

分子進化の数学モデルとして良く使われるのは Poisson process であるが、最近の分子時計の結果をみると必ずしも Poisson process で記述できるとは限らない。原因は同じ遺伝子の進化速度が系統ごとに違うためや、突然変異率が時間と共に変化するためなどいろいろある。そこで、Compound poisson, Doubly stochastic point process, Time dependent renewal process などが使われている。

いろいろなデータが蓄積され、1968 年に木村資生氏により分子進化の中立説が提唱された。この時、木村氏は分子進化の一定性と速度が突然変異率によって決まる上限値を持っていることを主張した。また多型を作り出しているメカニズムと分子進化を起こしているメカニズムは異なったものではなく現象の違った側面であることを強調した。ここで多型というのは一つの集団中に 2 つまたはそれ以上の対立遺伝子が共存することを言う。

中立突然変異に対しては進化速度 (r) は突然変異 (v) 率に等しい。ただし、一般にはすべての突然変異が中立とは限らず、全体のうち f の割合のものが淘汰に中立で、残りの $1-f$ の割合のものは有害で、進化に寄与しないとすると、そうすると、総突然変異率 v_T と中立突然変異 v の間には $v = v_T f$ という関係が成り立つ。進化速度は固定確率を U とすると $r = 2NvU = \frac{2Nv}{2N} = v$ となる。

多型に関しては、集団内の多型を 2 つランダムに選び、その 2 つの間にどの位 DNA 又は塩基座位に違いがあるかを調べる。任意交配集団において 2 個の遺伝子をランダムに選んだ時、共通の祖先に達するまでにかかる平均時間は $2N$ 世代である。これが複数個の遺伝子をランダムに選ぶと、かかる平均時間は $4N$ 世代になる。(図 9) よって $4N$ 世代の内に生じる違いは DNA 当たりの突然変異率が v ならば $\Pi = 4Nv$ である。また塩基座位当たりの違いは塩基座位当たりの突然変異率が μ ならば違いは $\pi = 4N\mu$ である。実際にデータ解析でも同じことをしており、いろいろな地域のヒトの遺伝子を選び、どのような系統関係があるかを調べたものが図 10 である。また 2 つの個体から遺伝子を選んだ場合が図 11 である。図 11 はヒトの 49 の遺伝子の異なった個体間でどの位違いがあるかを、DNA の機能的重要性が違う領域ごとに区切って多様性を調べた。平均的な DNA の違いは 0.1 % である。つまり 1000 塩基対当たり 1 個違うだけである。60 億人いるにも関わらず、こんなに遺伝的組成の違いが小さいということは、ヒトは小さい集団が最近急に大きくなったことを表わしている。

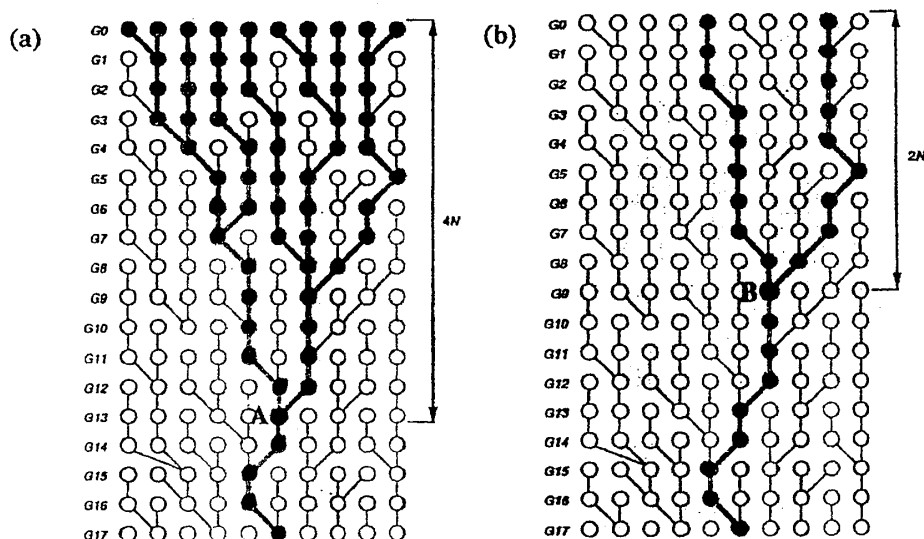


図 9 遺伝子の由来関係

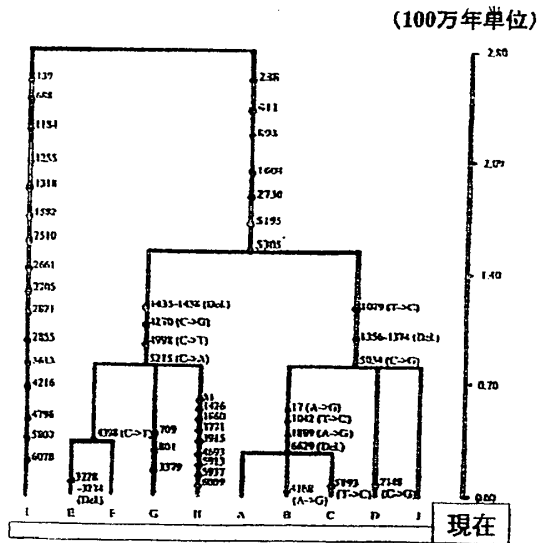


図10 遺伝子の系図

塩基多様度 π (%)

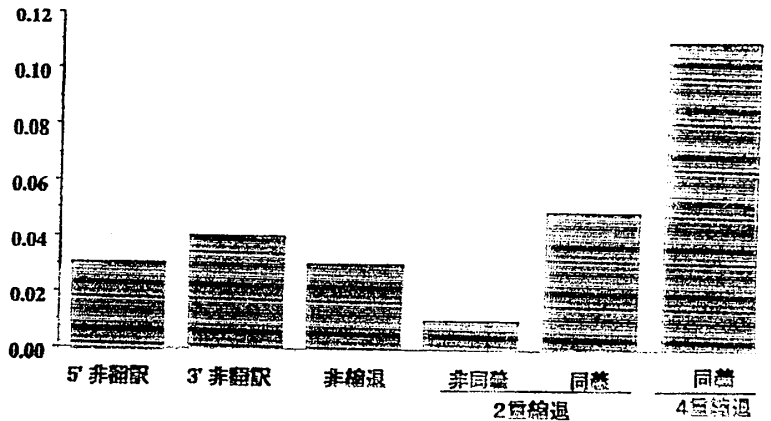


図11 ヒトの49の遺伝子座におけるDNA多型

N 個の遺伝子が共通の祖先に辿り着くまでの時間や過程は調べるには Death process が有効である。しかし、diffusion で調べるより実際のデータから作った tree と理論的な系図を合わせた方が見やすい。

§ 中立説の問題点

中立説のこれからの問題点として残っているものもある。そのうちの1つが、分子レベルの進化と表現型レベルの進化との間の際立った違いが存在することである。進化速度については、すでに述べたように分子レベルでは分子進化時計と呼ばれる性質があり、表現型的にはほとんど変わっていない生物でも、急速に進化してきた生物でもほとんど同じ速さで塩基の置換が起こる。また保守性がある。一方、表現型レベルの進化の特徴は便宜主義的進化である。つまり、環境に適応したものが多く子孫を残し、適応出来ないものは死滅する。ヒトとチンパンジーの遺伝子の間には1~2%しか違いがないのに表現型でみると大きな違いがある(図12)。表現型レベルの大きな違いの原因はまだ現在解明されていない。

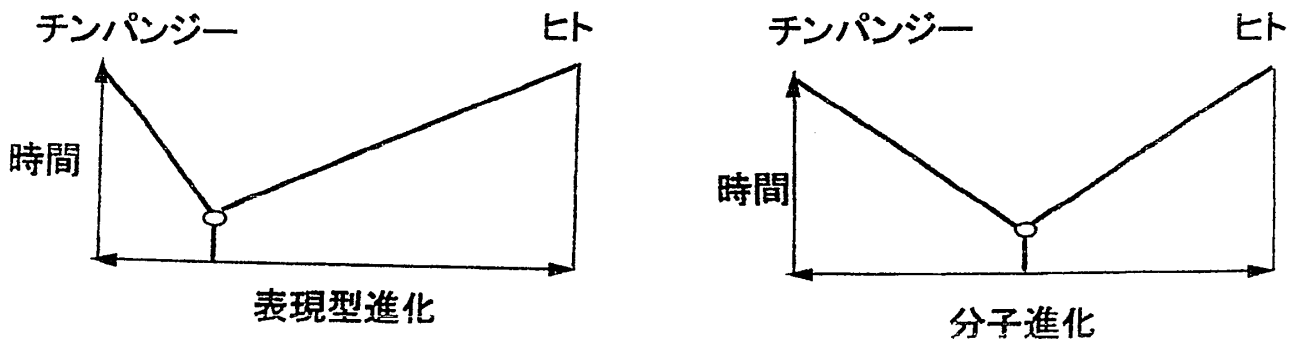


図12 分子進化と表現型進化の関係